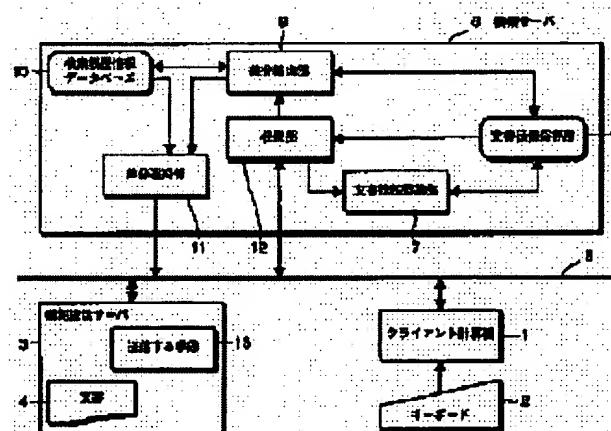


DEVICE FOR RETRIEVING DATA AND METHOD THEREFOR, AND INFORMATION PROCESSING SYSTEM AND METHOD THEREFOR, AND RECORDING MEDIUM FOR RECORDING PROGRAM FOR RETRIEVING DATA

Patent number: JP11110410
Publication date: 1999-04-23
Inventor: IMAMURA DAISUKE
Applicant: TOKYO SHIBAURA ELECTRIC CO
Classification:
- international: G06F17/30
- european:
Application number: JP19970275656 19971008
Priority number(s): JP19970275656 19971008

Report a data error here**Abstract of JP11110410**

PROBLEM TO BE SOLVED: To provide a data retrieving device for effectively feedbacking the taste of a user to an information provider side. **SOLUTION:** Document information related with a document 4 provided by an information providing server 3 is stored in a document information preserving part 8 of a retrieval server 6. A user designates a retrieval word from a client calculator 1, and requests retrieval to the retrieval server 6. A retrieving part 12 retrieves document information matched with the retrieval keyword, and answers the hit information providing server 3. A difference extracting part 9 extracts a difference between the stored document information and the retrieval keyword used by the user, and a difference communicating part 11 provides the information of the difference to the information providing server 3. The information provider captures the taste of the user from the information related with the difference, and matches information to be provided with the taste of the user for increasing the reading of a document.



Data supplied from the esp@cenet database - Worldwide

(19)日本国特許庁 (JP)

(12) 公開特許公報 (A)

(11)特許出願公開番号

特開平11-110410

(43)公開日 平成11年(1999)4月23日

(51)Int.Cl.⁶
G 0 6 F 17/30

識別記号

F I
G 0 6 F 15/403

3 5 0 C
3 4 0 A

審査請求 未請求 請求項の数13 OL (全 12 頁)

(21)出願番号 特願平9-275656

(22)出願日 平成9年(1997)10月8日

(71)出願人 000003078

株式会社東芝
神奈川県川崎市幸区堀川町72番地

(72)発明者 今村 大輔

東京都府中市東芝町1番地 株式会社東芝
府中工場内

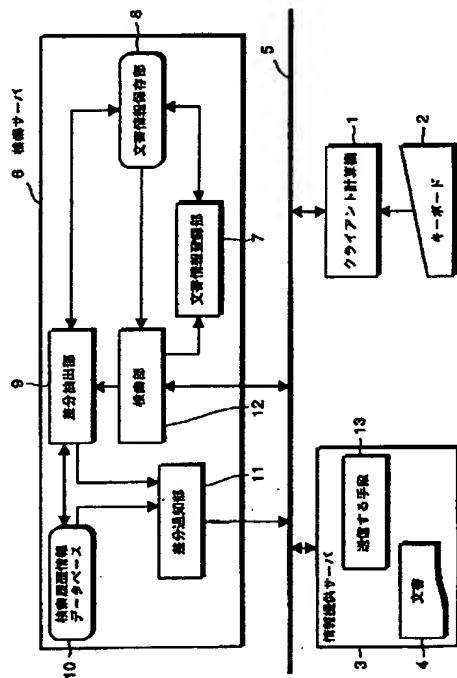
(74)代理人 弁理士 木内 光春

(54)【発明の名称】 データ検索装置及び方法、情報処理システム及び方法並びにデータ検索用プログラムを記録した記録媒体

(57)【要約】

【課題】 ユーザの嗜好を情報提供元側に効果的にフィードバックするデータ検索装置を提供する。

【解決手段】 情報提供サーバ3の提供する文書4に関する文書情報を、検索サーバ6の文書情報保存部8に蓄積する。ユーザは、クライアント計算機1から検索キーワードを指定して検索サーバ6に検索を要求する。検索部12は検索キーワードに合致する文書情報を検索し、ヒットした情報提供サーバ3を回答する。差分抽出部9は、蓄積されていた文書情報とユーザの用いた検索キーワードとの差分を抽出し、差分通知部11が差分の情報を情報提供サーバ3へ提供する。情報提供元は、このような差分に関する情報からユーザの嗜好を把握し、提供する情報をユーザの嗜好に合致させることによって、文書の閲覧を増やすことができる。



【特許請求の範囲】

【請求項1】 情報提供元から提供される情報に関する検索用情報を蓄積する手段と、前記検索用情報に基づいて、与えられた検索キーワードと合致する情報の提供元を検索する手段とを有するデータ検索装置において、前記検索用情報と検索キーワードとの差分を抽出する手段と、抽出された差分を、前記情報提供元に提供する手段と、を有することを特徴とするデータ検索装置。

【請求項2】 情報提供元から提供される情報について少なくとも当該情報に含まれるキーワードを検索用情報として蓄積する手段と、前記検索用情報に基づいて、与えられた検索キーワードと合致する情報の提供元を検索する手段とを有するデータ検索装置において、前記各情報提供元ごとに、検索の履歴を記録する手段と、

前記検索の履歴に基づいて、キーワードごとの重要度を計算する手段と、前記検索用情報と検索キーワードとの差分を抽出する手段と、抽出された差分と前記キーワードごとの重要度を、前記情報提供元に提供する手段と、を有することを特徴とするデータ検索装置。

【請求項3】 前記検索用情報に含まれないが重要度が高いキーワードについて、情報提供元へ所定の通知を行う手段を有することを特徴とする請求項2記載のデータ検索装置。

【請求項4】 前記通知が行われたキーワードについて、所定の公告を行う手段を有することを特徴とする請求項3記載のデータ検索装置。

【請求項5】 前記公告に係るキーワードが、情報提供元から提供される新たな情報に含まれる場合に、公告の内容を更新する手段を有することを特徴とする請求項4記載のデータ検索装置。

【請求項6】 前記検索用情報に含まれないが重要度が高いキーワードを、前記検索用情報に追加する手段を有することを特徴とする請求項1乃至5のいずれか一に記載のデータ検索装置。

【請求項7】 前記検索用情報に含まれているが重要度が低いキーワードを、前記検索用情報から削除する手段を有することを特徴とする請求項1乃至6のいずれか一に記載のデータ検索装置。

【請求項8】 情報提供元から提供される情報に関する検索用情報を蓄積し、前記検索用情報に基づいて、与えられた検索キーワードと合致する情報の提供元を検索するデータ検索方法において、

前記検索用情報と検索キーワードとの差分を抽出して、抽出された差分を、前記情報提供元に提供することを特徴とするデータ検索方法。

【請求項9】 情報提供元から提供される情報について少なくとも当該情報に含まれるキーワードを検索用情報として蓄積するステップと、前記検索用情報に基づいて、与えられた検索キーワードと合致する情報の提供元を検索するステップとを有するデータ検索方法において、

前記各情報提供元ごとに、検索の履歴を記録するステップと、

前記検索の履歴に基づいて、キーワードごとの重要度を計算するステップと、

前記検索用情報と検索キーワードとの差分を抽出するステップと、

抽出された差分とキーワードごとの重要度を、前記情報提供元に提供するステップと、を含むことを特徴とするデータ検索方法。

【請求項10】 情報を提供する第1のサーバと、前記第1のサーバが提供する情報に関する検索用情報を蓄積し、与えられた検索キーワードに基づいて前記検索用情報を検索する第2のサーバとがコンピュータネットワークで接続された情報処理システムにおいて、前記第2のサーバは、前記検索用情報に含まれていないが重要性が高いと判断されるキーワードを抽出する手段と、

抽出されたキーワードを前記第1のサーバへ通知する手段を有し、

前記第1のサーバは、新たに提供する情報に関する新たな検索用情報を、前記第2のサーバへ送信する手段を有することを特徴とする情報処理システム。

【請求項11】 情報を提供する第1のサーバと、前記第1のサーバが提供する情報に関する検索用情報を蓄積し、与えられた検索キーワードに基づいて前記検索用情報を検索する第2のサーバとがコンピュータネットワークで接続された情報処理方法において、

前記第2のサーバは、前記検索用情報に含まれていないが重要性が高いと判断されるキーワードを抽出し、抽出されたキーワードを前記第1のサーバへ通知し、

前記第1のサーバは、新たに提供する情報に関する新たな検索用情報を、前記第2のサーバへ送信することを特徴とする情報処理方法。

【請求項12】 情報提供元から提供される情報に関する検索用情報を蓄積させるステップと、前記検索用情報に基づいて、与えられた検索キーワードと合致する情報の提供元を検索させるステップとを有するデータ検索用プログラムを記録した記録媒体において、

前記プログラムが、

前記検索用情報と検索キーワードとの差分を抽出させるステップと、

抽出された差分を、前記情報提供元に提供させるステップを含むことを特徴とするデータ検索用プログラムを記録した記録媒体。

【請求項13】 情報提供元から提供される情報に関する検索用情報を蓄積させるステップと、前記検索用情報に基づいて、与えられた検索キーワードと合致する情報の提供元を検索させるステップと有するデータ検索用プログラムを記録した記録媒体において、
前記各情報提供元ごとに、検索の履歴を記録させるステップと、
前記検索の履歴に基づいて、キーワードごとの重要度を計算させるステップと、
前記検索用情報と検索キーワードとの差分を抽出させるステップと、
抽出された差分とキーワードごとの重要度を、前記情報提供元に提供させるステップと、
を含むことを特徴とするデータ検索用プログラムを記録した記録媒体。

【発明の詳細な説明】**【0001】**

【発明の属する技術分野】 本発明は、データベースをキーワード検索する技術の改良に関するもので、ユーザの嗜好を情報提供元側に効果的にフィードバックするようにしたものである。

【0002】

【従来の技術】 近年、コンピュータを用いたデータ検索装置が、いろいろな分野で普及している。例えば、インターネットでは、情報提供元となる多数のサーバが存在し、WWW(World Wide Web)上のホームページ、ニュースグループ、メーリングリストなどの情報が多数のサーバに分散している。このようなインターネットにデータ検索装置を適用したものは検索サーバ、検索エンジン、サーチエンジンなどと呼ばれる。

【0003】 インターネットの検索サーバでは、どのようなキーワードを含む情報が、どのサーバに存在するかというデータを予め蓄積しておく。このようなデータを文書情報と呼ぶ。文書情報の例としては、持っている情報のキーワードをサーバごとに列挙したものが考えられるが、より詳しい文書情報としては、あるサーバの持っている多数のホームページのうち、どのホームページがどのキーワードを含むか、のようなデータも考えられる。

【0004】 また、検索サーバが文書情報を蓄積するには、情報提供元が自発的に行う登録要請を受け付けるようにもよいし、専用のプログラムが情報提供元に巡回アクセスすることによって情報を収集するようにしてよい。また、文書情報として蓄積するキーワードは、情報提供元が決定してもよいし、検索サーバ側で決定してもよい。

【0005】 文書情報を蓄積した検索サーバに対してユーザが検索を要求するときは、キーワードを指定する。このように検索の際に指定するキーワードは「検索キーワード」と呼び、一般的には、複数のキーワードを AND

ORなどの論理演算子で組み合わせた検索式の形で入力する。このような検索式を与えられた検索サーバは、検索式に合致(ヒット)する文書情報を検索し、その文書情報に対応するサーバのネットワークアドレスや、さらにホームページの名称などを回答する。

【0006】

【発明が解決しようとする課題】 ところで、検索でヒットしたサーバにユーザがアクセスしても、真にユーザが欲していた情報が掲載されていることは希で、ユーザの嗜好と、サーバの提供する情報を合致させることが望まれている。しかしながら、上記のような従来技術は、サーバに関する検索結果をユーザが一方的に受け取るだけで、ユーザの嗜好を情報提供元に反映させるものではない。検索結果を出力する際、ヒットした文書が検索キーワードをどのくらい含むかの含有率を得点として表示する検索サーバも存在するが、これもユーザにとってはあくまでも受動的に情報を受け取るもので、自らの嗜好を情報提供元側に伝達するものではない。また、情報提供元にあっても、ユーザの嗜好を把握するためには、アンケートなどの手段をあらためて講ずる必要があった。

【0007】 本発明は、上記のような従来技術の問題点を解決するために提案されたもので、その第1の目的は、ユーザの嗜好を情報提供元側に効果的にフィードバックすることができる情報検索装置及び情報検索方法を提供することにある。

【0008】 また、本発明の第2の目的は、ユーザの嗜好を情報提供元側に効果的にフィードバックし、情報提供元からユーザのニーズに合致した情報を提供できるようにした効率の良い情報処理システム及び情報処理方法を提供することにある。

【0009】

【課題を解決するための手段】 上記の目的を達成するため、請求項1に記載の発明は、情報提供元から提供される情報に関する検索用情報を蓄積する手段と、前記検索用情報に基づいて、与えられた検索キーワードと合致する情報の提供元を検索する手段とを有するデータ検索装置において、前記検索用情報と検索キーワードとの差分を抽出する手段と、抽出された差分を、前記情報提供元に提供する手段とを有することを特徴とするものである。

【0010】 また、請求項8に記載の発明は、請求項1に記載の発明を方法の観点から捉えたものであって、情報提供元から提供される情報に関する検索用情報を蓄積し、前記検索用情報に基づいて、与えられた検索キーワードと合致する情報の提供元を検索するデータ検索方法において、前記検索用情報と検索キーワードとの差分を抽出し、抽出された差分を前記情報提供元に提供することを特徴とするものである。

【0011】 さらに、請求項12の発明は、請求項8に記載の発明を、コンピュータプログラムを記録した記録

媒体の観点から捉えたものであって、情報提供元から提供される情報に関する検索用情報を蓄積させるステップと、前記検索用情報に基づいて、与えられた検索キーワードと合致する情報の提供元を検索させるステップを有するデータ検索用プログラムを記録した記録媒体において、前記プログラムが、前記検索用情報と検索キーワードとの差分を抽出させるステップと、抽出された差分を、前記情報提供元に提供させるステップを含むことを特徴とするものである。

【0012】上記のような構成を有する請求項1, 8, 12に記載された発明では、情報提供元が、検索用情報と検索キーワードとの差分に基づいて、提供している情報とユーザの嗜好とのずれを判断することができるので、ユーザの嗜好が情報提供元へ効果的にフィードバックされる。

【0013】請求項2に記載の発明は、情報提供元から提供される情報について少なくとも当該情報に含まれるキーワードを検索用情報として蓄積する手段と、前記検索用情報に基づいて、与えられた検索キーワードと合致する情報の提供元を検索する手段とを有するデータ検索装置において、前記各情報提供元ごとに、検索の履歴を記録する手段と、前記検索の履歴に基づいて、キーワードごとの重要度を計算する手段と、前記検索用情報と検索キーワードとの差分を抽出する手段と、抽出された差分と前記キーワードごとの重要度を、前記情報提供元に提供する手段とを有することを特徴とするものである。

【0014】また、請求項9に記載の発明は、請求項2に記載の発明を方法の観点から捉えたものであって、情報提供元から提供される情報について少なくとも当該情報に含まれるキーワードを検索用情報として蓄積するステップと、前記検索用情報に基づいて、与えられた検索キーワードと合致する情報の提供元を検索するステップとを有するデータ検索方法において、前記各情報提供元ごとに、検索の履歴を記録する手段と、前記検索の履歴に基づいて、キーワードごとの重要度を計算する手段と、前記検索用情報と検索キーワードとの差分を抽出するステップと、抽出された差分とキーワードごとの重要度を、前記情報提供元に提供する手段とを含むことを特徴とするものである。

【0015】さらに、請求項13の発明は、請求項9に記載の発明を、コンピュータプログラムを記録した記録媒体の観点から捉えたものであって、情報提供元から提供される情報に関する検索用情報を蓄積させるステップと、前記検索用情報に基づいて、与えられた検索キーワードと合致する情報の提供元を検索させるステップを有するデータ検索用プログラムを記録した記録媒体において、前記各情報提供元ごとに、検索の履歴を記録させるステップと、前記検索の履歴に基づいて、キーワードごとの重要度を計算させるステップと、前記検索用情報と検索キーワードとの差分を抽出させるステップと、抽出

された差分とキーワードごとの重要度を、前記情報提供元に提供させるステップとを含むことを特徴とするものである。

【0016】上記のような構成を有する請求項2, 9, 13に記載の発明では、情報提供元は、検索の履歴から計算したキーワードごとの重要度に基づいて、ユーザの嗜好をより詳細に理解することができる。

【0017】請求項3記載の発明は、請求項2に記載のデータ検索装置において、前記検索用情報に含まれないが重要度が高いキーワードについて情報提供元へ所定の通知を行う手段を有することを特徴とするものである。このような構成を有する請求項3記載の発明では、検索用情報に含まれないが重要度が高いキーワード、例えば、検索キーワードとして繰り返し指定されるようなキーワードについては、情報提供元へ所定の通知が行われるので、情報提供元はユーザの嗜好に合った新たな情報を提供することが容易になる。

【0018】請求項4記載の発明は、請求項3記載のデータ検索装置において、前記通知が行われたキーワードについて、所定の公告を行う手段を有することを特徴とするものである。このような構成を有する請求項4記載の発明では、通知が行われたキーワードについては、情報提供元からそのキーワードを含む新たな情報が提供される可能性が高く、また、そのキーワードが予め公告されるので、ユーザは新たな情報が提供される可能性を早期に知ることができ、提供される新たな情報の利用率が改善される。

【0019】請求項5記載の発明は、請求項4記載のデータ検索装置において、前記公告に係るキーワードが、情報提供元から提供される新たな情報に含まれる場合に、公告の内容を更新する手段を有することを特徴とするものである。このような構成を有する請求項5記載の発明では、公告されていたキーワードを含む文書が実際に追加された場合に、その旨を反映して公告内容が更新されるので、ユーザは新たな文書の追加を知ることができ、追加された文書がユーザによって利用される可能性が向上する。

【0020】請求項6記載の発明は、請求項1乃至5のいずれか一に記載のデータ検索装置において、前記検索用情報に含まれないが重要度が高いキーワードを前記検索用情報に追加する手段を有することを特徴とするものである。このような構成を有する請求項6記載の発明では、検索用情報に含まれないが、重要度が高いキーワードが検索用情報に追加されるので、検索におけるヒット率が改善される。

【0021】請求項7の発明は、請求項1乃至6のいずれか一に記載のデータ検索装置において、重要度が低いキーワードを検索用情報から削除する手段を有することを特徴とするものである。このような構成を有する請求項7記載の発明では、重要度が低いキーワード、例え

ば、検索キーワードとして用いられないようなキーワードが検索用情報から削除されるので、無駄なキーワードが減少し、検索処理が効率化される。

【0022】請求項10に記載の発明は、情報を提供する第1のサーバと、前記第1のサーバが提供する情報に関する検索用情報を蓄積し、与えられた検索キーワードに基づいて前記検索用情報を検索する第2のサーバとがコンピュータネットワークで接続された情報処理システムにおいて、前記第2のサーバは、前記検索用情報に含まれていないが重要性が高いと判断されるキーワードを抽出する手段と、抽出されたキーワードを前記第1のサーバへ通知する手段を有し、前記第1のサーバは、新たに提供する情報に関する新たな検索用情報を、前記第2のサーバへ送信する手段を有することを特徴とするものである。

【0023】また、請求項11に記載の発明は、上記請求項10に記載の発明を方法の観点から捉えたものであって、情報を提供する第1のサーバと、前記第1のサーバが提供する情報に関する検索用情報を蓄積し、与えられた検索キーワードに基づいて前記検索用情報を検索する第2のサーバとがコンピュータネットワークで接続された情報処理方法において、前記第2のサーバは、前記検索用情報に含まれていないが重要性が高いと判断されるキーワードを抽出し、抽出されたキーワードを前記第1のサーバへ通知し、前記第1のサーバは、新たに提供する情報に関する新たな検索用情報を、前記第2のサーバへ送信することを特徴とするものである。情報処理方法。

【0024】上記のような構成を有する請求項10, 11記載の発明では、検索用情報に含まれていないが検索の履歴から見て重要性が高いと判断されるキーワードが、情報を提供する第1のサーバへ通知される。これにより、第1のサーバは、通知されたキーワードを含む新たな情報を提供することができる。また、この新たな情報はユーザーの嗜好に合致した情報であるので、この新たな情報に関する検索用情報を、第2のサーバへ送信しておけば、ユーザーによる利用が促進される。

【0025】

【発明の実施の形態】以下、本発明の情報処理システムの実施の形態（以下、「本実施形態」という）について、図面に基づいて具体的に説明する。なお、本実施形態は、コンピュータのCPU及び周辺機器をプログラムが制御することによって実現され、ハードウェアやソフトウェアの具体的な構成は種々考えられる。このため、以下では、本実施形態の各機能に対応する「～部」などの仮想的回路ブロックを用いて説明する。なお、CPUは、プログラムで指定されたとおりに、コンピュータの各種ハードウェア資源を利用しながら、前記各仮想的回路ブロックの作用を実現する。

【0026】CPUには、ハードウェア資源の典型例と

して、バス及び入出力制御回路を介して、RAMなどの記憶素子からなるメモリ、ハードディスクドライブなどの補助記憶装置、入力装置としてマウスやキーボード、出力装置として表示装置やプリンタを接続することが考えられる。但し、これらハードウェア資源は例示に過ぎず、情報の記憶・入力・出力などの目的を達成できる他の各種装置を用いることもできる。

【0027】〔1. 構成〕図1は、本実施形態の構成を示す機能ブロック図である。本実施形態は、この図に示すように、クライアント計算機1、情報提供サーバ3及び検索サーバ6をネットワーク5を介して接続したもので、どの情報提供サーバ3がどのような情報を提供するかを表す文書情報を検索サーバ6に蓄積し、ユーザーはクライアント計算機1から検索サーバ6にアクセスすることによって、自分の欲しい情報をどの情報提供サーバ3が持っているか検索するものである。

【0028】このうちクライアント計算機1は、ネットワーク5に接続された情報端末であり、ユーザーが、検索キーワードを検索式の形で入力することによって検索サーバ6に検索を要求したり、検索で判明した情報提供サーバ3にアクセスすることによって目的の情報を表示するのに用いる。このクライアント計算機1には、文字列を入力するためのキーボード2など、必要な周辺装置が接続されている。

【0029】次に、情報提供サーバ3は、情報として文書4を提供する情報提供元となる計算機システムであり、クライアント計算機1など外部から情報の提供を求められたときに、文書4の内容を返送することによって公開する。また、この情報提供サーバ3は、また、検索サーバ6に対しては、文書4に含まれているキーワードなどの文書情報を提供する。なお、説明を単純にするために、ここでは、情報提供サーバ3ごとに一種類の文書情報が存在するものとする。

【0030】なお、情報提供サーバ3は、図1では1つだけを示すが、ネットワーク5には同様の情報提供サーバ3が多数接続され、それぞれの情報提供サーバ3が様々な文書4を提供するものとする。ネットワーク5は、TCP/IPなどの通信プロトコルに則ったデータ転送を実現する設備で、有線回線又は無線回線のいずれを用いて実現してもよい。

【0031】検索サーバ6は、クライアント計算機1から検索要求を受けたときに、蓄積してある文書情報のなかから検索式にマッチするものを検索することによって、ユーザーが欲しい情報がどの情報提供サーバ3に存在するかを回答するサーバである。

【0032】この検索サーバ6は、まず、文書情報登録部7及び文書情報保存部8を有している。このうち文書情報登録部7は、それぞれの情報提供サーバ3に関する文書情報を文書情報保存部8に登録する手段である。また、文書情報保存部8は、各情報提供サーバごとに、そ

のサーバによって提供される文書4に含まれる少なくともキーワード等の文書情報と、その文書情報には含まれていないが、ユーザからの検索要求が多いキーワードであって、新たなキーワードとして追加されたキーワード（以下、「追加キーワード」という）を保存する部分である。文書情報保存部8に保存される情報のうち、登録された文書情報の例を表1に示す。

【0033】

【表1】

情報提供サーバ名	文書情報
サーバ1	KeyA KeyB KeyC
サーバ2	KeyB, KeyC
サーバ3	KeyA, KeyB, KeyC, KeyD
:	:

【0034】この例では、情報提供サーバごとの情報として、そのサーバが提供する文書に含まれるキーワードが列挙されている。また、情報提供サーバ名はネットワーク上での識別名称であり、表1では仮にサーバ1, 2, 3としてある。

【0035】また、文書情報保存部8に保存される情報のうち、追加キーワードの例を表2に示す。

【0036】

【表2】

情報提供サーバ名	追加キーワード
サーバ1	KeyD
	KeyF
	KeyG
サーバ2	KeyA
	KeyF
サーバ3	KeyC
	KeyF
:	:

	検索キーワード	訪問実績	不足キーワード	過多キーワード
履歴1	KeyA, KeyB, KeyC	1	KeyD	KeyC
履歴2	KeyA, KeyC, KeyD	0	KeyB	KeyB
履歴3	KeyB, KeyD	1	KeyD	KeyA, KeyC
履歴4	KeyA, KeyB, KeyD	1	KeyD	KeyC

差分情報	-KeyA-2KeyB-3KeyC+3KeyD+2KeyE

【0041】この例では、検索ごとに用いられた検索キーワードの組み合わせが「KeyA, KeyB, KeyC」のように保存されている。また、「不足キーワード」とび「過多キーワード」の欄は、差分抽出部9によって抽出された前記差分の具体的な内容である。さらに、「訪問実績」とは、0又は1のいずれかの値をとるフラグで、検索でその情報提供サーバ3を知ったユーザが、実際にアクセスして文書4を閲覧した場合に1となる。

【0037】この例では、情報提供サーバごとにいくつかの追加キーワードが保存されている。

【0038】また、検索サーバ6は、検索部12と、差分抽出部9と、検索履歴情報データベース10と、差分通知部11とを有する。このうち検索部12は、クライアント計算機1から送られてきた検索式にマッチする文書情報を文書情報保存部8から検索し、マッチした文書情報の情報提供サーバ3をクライアント計算機1に回答する部分である。

【0039】また、差分抽出部9は、受け付けた検索式と文書情報保存部8に保存されている内容を比較することによって、両者の差分を抽出する部分である。また、差分抽出部9は、後述する検索履歴情報データベース10の内容を、検索式と、抽出した差分によって更新する機能と、新たなキーワードを文書情報保存部8に追加キーワードとして登録する機能を持つ。

【0040】検索履歴情報データベース10は、情報提供サーバごとに、当該情報提供サーバにヒットした過去の検索履歴を格納するデータベースである。ここで、検索履歴情報データベース10に保存された情報の例を表3に示す。

【表3】

る。

【0042】また、「不足キーワード」とは、検索式には含まれていたが、ヒットした情報提供サーバ3の文書情報には含まれていなかったキーワードであり、言い換えば、ユーザが必要としているキーワードである。一方、「過多キーワード」とは、逆に、ヒットした情報提供サーバ3の文書情報には含まれていたが、検索式には含まれていなかったキーワードであり、言い換えば、

ユーザが必要としていないキーワードである。なお、表3最下部の「差分情報」は、差分であるこれら「不足キーワード」及び「過多キーワード」及び「訪問実績」に基づいて、所定の基準で計算される情報である。この計算方法については、後述する。

【0043】差分通知部11は、所定の場合に差分抽出部9からのトリガ（起動命令）によって起動され、検索履歴情報データベース10の内容を参照することによって、情報提供サーバ3に対して、キーワードを用いた検索の傾向について、所定の情報を提供する部分である。

【0044】また、情報提供サーバ3は、新たな文書4の提供を開始する際に、新たな文書情報を検索サーバ6に送信する手段13を有する。

【0045】〔2. 作用及び効果〕上記のような構成を有する本実施の形態は次のような作用を有する。まず、図2は、検索サーバ6が、クライアント計算機1から検索キーワードを受け付けた場合の処理手順を示すフローチャートである。

【0046】〔2-1. 検索キーワードを受け付けた場合の検索処理〕すなわち、ユーザは、自分の欲しい情報の検索キーワードから検索式を作成してクライアント計算機1に入力すると、ネットワークを通じて、入力した検索式が検索要求として検索サーバ6に送信される。検索サーバ6の検索部12は、この検索式を受け付けると（ステップ201）、文書情報保存部8に保存されている文書情報（表1）から検索式に合致（ヒット）するものを検索し、その文書情報に対応する情報提供サーバ名を、検索結果としてクライアント計算機1に提示する（ステップ202）。

【0047】例えば、表1に示す文書情報が文書情報保存部8に保存されているとして、クライアント計算機1から3つの検索キーワードKey A, Key B, Key Dが送られてきたとすると、表1の3つのサーバ1, 2, 3のうち、検索キーワードと一致の度合いが最も高い「サーバ1」がヒットした情報提供サーバとしてクライアント計算機1に提示される。この検索結果を見たユーザは、そのままクライアント計算機1を用いて「サーバ1」にアクセスすることによって所望の情報を閲覧することができる。

【0048】なお、検索部12は、検索結果を提示する際に、情報提供サーバ3毎に、文書情報と検索式の間で一致したキーワードの個数を得点として提示するように構成することもできる。検索式と、その検索でヒットした情報提供サーバ名は、検索部12から差分抽出部9へ渡される。

【0049】〔2-2. 差分の抽出と検索履歴の更新〕続いて、差分抽出部9は、ヒットした情報提供サーバの文書情報と検索式とを比較するなどして両者の差分を抽出し、検索式と、抽出した差分とを用いて検索履歴情報データベース10に保存されている検索履歴を更新する

（ステップ203）。差分は、不足キーワード及び過多キーワードである。抽出したこれらのキーワードは、個々の情報提供サーバ3ごとに、検索履歴と共に検索履歴情報データベース10に記録する。なお、検索式中のキーワードが、andやorといった論理演算子で結合されていたような場合は、個々のキーワードに分解したうえで、差分の抽出と検索履歴の更新を行う。

【0050】表3に示した検索履歴の例では、第1回目の検索による履歴1において、検索式に検索キーワードKey A, Key B, Key Dが含まれていたことが記録されている。また、履歴1の差分のうち、訪問実績が「1」になっており、これは、検索から一定時間内に、ユーザがこの情報提供サーバに実際にアクセスして情報の提供を受けたことを示している。この訪問実績を得るには、検索の後一定時間の間、クライアント計算機1とヒットした情報提供サーバ3のアクセス動作を監視し、その間に情報提供サーバ3とクライアント計算機1の間でデータのやりとりがあれば訪問実績のフラグを1にセットすればよい。

【0051】また、表3の履歴1では、不足キーワードとしてKey Dが記録されている。これは、ユーザが作成した検索式にはキーワードKey Dが含まれていたが、ヒットした「サーバ1」の文書情報にはこのキーワードKey Dが含まれていなかつたことを意味する。また、履歴1の過多キーワードとしてはKey Cが記録されており、これは、ヒットした「サーバ1」の文書情報にこのキーワードKey Cが含まれていたが、このキーワードKey Cは検索式に用いられていなかつたことを意味する。

【0052】〔2-3. 重要度の計算〕続いて、差分抽出部9は、検索履歴情報データベース10内の検索履歴に含まれる不足キーワード及び過多キーワードについて、重要度を計算するかどうかを判断する（ステップ204）。不足キーワード及び過多キーワードについては、文書情報保存部8に追加したり削除するかどうかの判断基準とするために得点を計算するが、この計算を検索毎ではなく、履歴の更新がある程度まとまった時点でまとめてバッチ処理的に計算を行うことによって、検索サーバ6の負荷を低減するものである。重要度を計算するタイミングは、例えば、前回計算した後で一定回数の検索がされたときや、一定個数の過不足キーワードが蓄積されたとき、などとすればよい。ステップ204において、重要度を計算しないと判断された場合は処理を終了するが、重要度を計算する場合は次のような処理を行う（ステップ205）。

【0053】すなわち、ステップ205で計算される重要度は、不足キーワードのそれぞれ及び過多キーワードのそれぞれについて、検索履歴に登場した回数に応じて計算される符号付きの数値である。この重要度は、次のようにして計算する。まず、不足キーワードには正（こ

こでは+1とする)の係数を付け、過多キーワードには負(ここでは-1とする)の係数を付ける。また、ユーザがアクセスしたために訪問実績のフラグが1になっている履歴では係数を前記+1又は-1のままとし、ユーザがアクセスしなかったために訪問実績のフラグが0になっている履歴では、係数に重みを付ける。そして、このような係数と重みを前提として重要度を計算する。

【0054】例えば、訪問実績がない場合の重みを2とした場合、訪問実績がない履歴の不足キーワードや過多キーワードは、訪問実績がある履歴と比べて、2倍に評価される。このように、訪問実績がない履歴の過不足キーワードの重要度を大きく評価することによって、不人気な情報提供サーバ3に関するキーワードの修正が促進される。

【0055】今、文書情報保存部8に登録されている「サーバ1」に対応する文書情報のキーワードが、表1に示すようにKey A, Key B, Key Cであり、この「サーバ1」の検索履歴として、表3に示すように履歴1から履歴4までが保存されているものとする。この検索履歴から「サーバ1」に関する重要度を計算する例を示す。

【0056】まず、履歴1については、不足キーワードKey Dは+1点、過多キーワードKey Cは-1点という重要度を得るが、これは+Key D-Key Cと表すことができる。同様に、履歴2から履歴4についてもそれぞれ重要度を求め、これらを集計すると、+Key D-Key C+2Key E-2Key B+Key D-Key A-Key C+Key D-Key Cとなり、この式をさらに集約すれば、-Key A-2Key B-3Key C+3Key D+2Key Eという計算結果が得られる(表3最下欄)。このように最終的に集約された計算結果を差分情報と呼ぶ。

【0057】このように計算された各キーワードの係数は、正の絶対値が大きいほどユーザによって要求される頻度が大きく、負の絶対値が大きいほどユーザによって要求される頻度が低いことを意味する。そして、各キーワードについてこれらの係数がそのキーワードの履歴における重要度となる。例えば、表2最下欄の計算結果では、Key Aの重要度は-1であり、Key Bの重要度は-2であるから、このサーバにとって、Key BはKey Aよりも重要度が低いと判断できる。

【0058】[2-4. 不足キーワードと過多キーワードの取り扱い] 続いて、差分抽出部9は、上記のように計算された重要度に応じて、不足キーワードは文書情報保存部8に追加し、過多キーワードは文書情報保存部8から削除する。すなわち、差分抽出部9は、計算された差分情報に含まれる全ての不足キーワード及び過多キーワードについて処理が済むまで(ステップ213)、次のような手順を繰り返す。

【0059】まず、差分情報中のキーワードを1つ選択

し(ステップ206)、選択したキーワードの重要度が正の閾値を超えているときは(ステップ207)、そのキーワードが文書情報保存部8に登録されていないことを確認したうえ(ステップ208)、そのキーワードを文書情報保存部8への追加候補として所定のワークエリアに記録しておく(ステップ209)。

【0060】選択したキーワードの重要度がステップ207において正の閾値を超えていないときは、そのキーワードの重要度が負の閾値を超えていかどうかを判断する(ステップ210)。このとき、キーワードの重要度が負の閾値を超えているときは(ステップ210)、そのキーワードが文書情報保存部8からまだ削除されていないことを確認したうえ(ステップ211)、そのキーワードを、文書情報保存部8からの削除候補として所定のワークエリアに記録しておく(ステップ212)。

【0061】なお、正の閾値を超えたキーワードが文書情報保存部8に既に存在した場合(ステップ208)、負の閾値を超えたキーワードが文書情報保存部8から既に削除されていた場合(ステップ211)及び、追加や削除の候補を記録した(ステップ209, 212)後は、ステップ206に戻って次のキーワードを選択する。

【0062】ステップ206から213の処理を全ての不足キーワード及び過多キーワードに対して繰り返すと、差分抽出部9はトリガを送ることによって差分通知部11を起動し、差分通知部11は次のような処理を行う。まず、ワークエリアに記録してある追加候補のキーワードについては、文書情報保存部8へ追加キーワードとして追加してよいかどうかを情報提供サーバ3に問い合わせ、ワークエリアに記録してある削除候補については、文書情報保存部8からキーワードを削除したうえで、情報提供サーバ3にその旨を通知する(ステップ215)。

【0063】ここで、追加要請や削除を情報提供サーバ3に通知する方法としては、情報提供サーバ3の管理者の電子メールアドレスを予め登録しておき、そのアドレスに電子メールを送信してもよいし、情報提供サーバ3側に通知を受け付けるための専用のプロセスを常駐させてもよい。また、差分通知部11においてホームページなどの形式で通知したい内容を公開しておき、情報提供サーバ3側がこのホームページに適宜アクセスするようにもよい。なお、表3に示す不足キーワード、過多キーワード及び差分情報は、情報提供サーバ3から、必要に応じて読み出せるものとする。

【0064】ステップ215の問合せに対する返答(ステップ216)は、例えば、追加を認めることができない一部の追加候補を除いて、追加を認める、などの内容である。返答を受けた検索サーバ6は、追加が承諾された追加候補のキーワードを文書情報保存部8に追加キー

ワードとして追加する。ところで、情報提供サーバ3に追加の是非が問合せられたキーワードは、情報提供サーバ3がそのキーワードを含む文書を提供すれば、ユーザーの訪問（アクセス）をより多く得られることを意味するので、情報提供サーバ3はそのキーワードを含む文書4を近く提供する可能性が高い。このため、検索サーバ6においては、新しい追加キーワードについては、予告情報の形でユーザーに公開する（ステップ214）。

【0065】この予告情報は、そのキーワードを含む文書が情報提供サーバ3から近日中に公開される予定があるという意味の情報である。なお、予告情報の表示例を図3に示す。この表示例において、文字列501はキーワードを示し、文字列502は、キーワードを含む文書を公開する予定の情報提供サーバ3のインターネットアドレスを示している。

【0066】〔2-5. 追加キーワードの取り扱い〕次に、ステップ217で文書情報保存部8に追加された追加キーワードがその後どのように取り扱われるかについて説明する。すなわち、追加キーワードについては、上述したように、情報提供サーバによってその追加キーワードを含む情報が新たに提供される可能性が高い。従って、再度この追加キーワードを含む検索式が与えられた場合、ユーザーが欲する情報が得られる可能性が高くなり、検索におけるヒット率が改善される。

【0067】また、追加キーワードを含む新たな文書情報が正式に登録された場合に、追加キーワードが文書情報保存部8から削除される手順（図4）を説明する。

【0068】この手順では、まず、文書情報登録部7が、新たな文書情報を獲得する（ステップ401）。このように新たな文書情報を獲得するには、次の2つの態様が考えられる。一つは、情報提供元から自発的に新たな文書情報の登録申請を行い、これを検索サーバ6の側で受け付けることである。もう一つは、検索サーバ6の側から予告情報に述べられている情報提供サーバ3へ、検索ロボットやプローブなどと呼ばれる専用の情報収集用プログラムを派遣することである。検索ロボットは、派遣された先の情報提供サーバ3で新たに公開されている文書の内容を収集して検索サーバ6に持ち帰る。

【0069】これらのうちいずれかの態様で新たな文書情報が獲得されると、文書情報登録部7は、追加キーワードが新たな文書情報に含まれるかどうかを判断する（ステップ402）。追加キーワードが新たな文書情報に含まれる場合、文書情報登録部7は、その含まれる追加キーワードを、文書情報保存部8の追加キーワードから削除するとともに（ステップ403）、予告情報から新着情報に移動する（ステップ404）。

【0070】ここで、新着情報の表示例を図5に示す。新着情報は、予告情報と同様にユーザーへの公告の一種であるが、文書情報に正式に追加されたキーワードに関するものである。なお、ステップ402において、追加キ

ワードが新たな文書情報に含まれていない場合は、未だ、ユーザーの要望は達成されていないので、追加キーワードの削除（ステップ403）、予告情報から新着情報への移動（ステップ404）の処理は行わない。

【0071】続いて、文書情報登録部7は、文書情報保存部8に蓄積されている文書情報を新たな文書情報によって更新する（ステップ405）。また、文書情報登録部7は、続いて、追加キーワードとして追加された後、所定の日数を経過した追加キーワードがあるかどうかを判断し（ステップ406）、そのような追加キーワードがあった場合は、追加の予告が達成されなかったものとして、そのキーワードを文書情報保存部8から削除し（ステップ407）、予告情報からも削除する（ステップ408）。

【0072】〔3. 効果〕以上説明したように、本実施形態では、情報提供元が、検索用情報と検索キーワードとの差分に基づいて、提供している情報とユーザーの嗜好とのずれを判断することができるので、ユーザーの嗜好が情報提供元へ効果的にフィードバックされる。また、本実施形態では、情報提供元は、検索の履歴から計算したキーワードごとの重要度に基づいて、ユーザーの嗜好をより詳細に理解することができる。

【0073】また、本実施形態では、検索用情報に含まれないが重要度が高いキーワード、例えば、検索キーワードとして繰り返し指定されるようなキーワードについては、情報提供元へ所定の通知が行われるので、情報提供元はユーザーの嗜好に合った新たな情報を公開することが容易になる。また、このような通知が行われたキーワードについては、情報提供元からそのキーワードを含む新たな情報が提供される可能性が高いが、本実施形態では、そのキーワードが予め公告されるので、ユーザーは新たな情報が提供される可能性を早期に知ることができ、提供される新たな情報の利用率が改善される。そして、この新たな情報については、情報提供サーバ3から検索サーバ6へ新たな文書情報を送信しておけば、検索結果としてその情報提供サーバ3が提示されることになるので、新たな情報のユーザーによる利用が促進される。

【0074】また、本実施形態では、公告されていたキーワードを含む文書が実際に追加された場合に、その旨を反映して公告内容が更新されるので、ユーザーは新たな文書の追加を知ることができ、追加された文書がユーザーによって利用される可能性が向上する。

【0075】また、本実施形態では、文書情報には含まれないが、重要度が高いキーワードは追加情報として文書情報保存部8に追加されるので、検索におけるヒット率が改善される。また、本実施形態では、重要度が低いキーワード、すなわち、検索キーワードとして用いられないようなキーワードが検索用情報から削除されるので、無駄なキーワードが減少し、検索処理が効率化される。

【0076】以上のような本実施形態によれば、まず、広告などの文書を公開している情報提供元は、差分や重要度などの情報からユーザーの嗜好を把握し、提供する情報をユーザーの嗜好に合致させることによって、文書の閲覧を増やすことができる。また、不足キーワードや過多キーワードの形でユーザーのニーズを把握することができるので、提供する情報やその他の活動について、市場への適応が容易になる。

【0077】また、ユーザーにとっても、自らの嗜好に合った情報が増加するだけでなく、検索のヒット率が向上することによって所望の文書を発見・利用することが容易になる。さらに、データ検索装置の運営者も、検索のヒット率を向上させることによってシステムの利用効率や利用者数を増大させることができる。特に、情報提供者とユーザーの双方にとって検索の有用性が向上するため、ユーザーが検索を利用する頻度が増加し、情報提供者が提供する広告などの情報がユーザーに閲覧される頻度も向上する。

【0078】〔4. 他の実施の形態〕本発明は、上記の実施の形態に限定されるものではなく、次に例示するような他の実施の形態をも包含するものである。例えば、検索サーバ6が文書情報を蓄積する態様は自由で、例えば、情報提供元からの申請を受理する形でもよいし、自ら情報提供元へアクセスして情報を収集してもよいし、第三者から情報提供を受けてもよい。また、文書情報の具体的な内容は自由に定めることができ、キーワードの他に、例えばジャンルや最後のホームページが更新された日など、所望の情報を加えることができる。また、クライアント計算機1は複数設けてもよい。

【0079】また、上記実施形態では不足キーワードと過多キーワードの双方を抽出し、情報提供元に通知したが、このうち一方のみを抽出して提供者に通知するようにしてもよい。また、抽出された差分や重要度などの情報を情報提供元へ通知する態様も自由で、差分通知部11の側から情報提供サーバ3に対して差分情報を通知してもよいし、情報提供サーバ3から差分通知部11にアクセスすることによって差分情報を読み出してもよい。また、本発明はデータ検索用プログラムを用いて実現さ

れることが一般的と考えられるが、そのようなプログラムを記録した記録媒体も本発明の一態様である。

【0080】

【発明の効果】以上のように、本発明によれば、ユーザーの嗜好を情報提供元側に効果的にフィードバックすることができる情報検索装置及び情報検索方法を提供することができる。また、ユーザーの嗜好を情報提供元側に効果的にフィードバックし、情報提供元からユーザーのニーズに合致した情報を提供できるようにした効率の良い情報処理システム及び情報処理方法を提供することができる。

【図面の簡単な説明】

【図1】本発明の一実施形態の構成を示す機能ブロック図。

【図2】本発明の実施の形態において、情報提供サーバへの差分の通知までの処理手順を示すフローチャート。

【図3】本発明の実施の形態における予告情報の表示例を示す図。

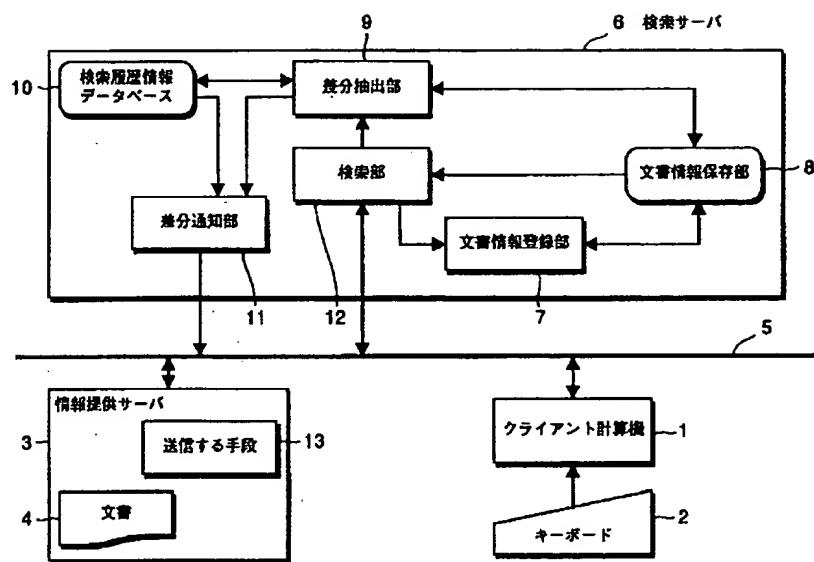
【図4】本発明の実施の形態において、新たな文書情報の登録によって追加キーワードが文書情報保存部から削除される手順を示すフローチャート。

【図5】本発明の実施の形態における新着情報の表示例を示す図。

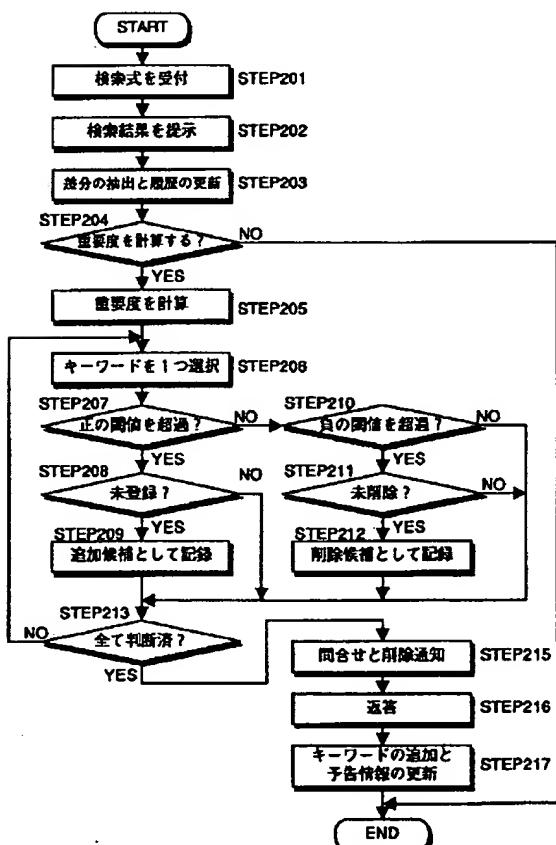
【符号の説明】

- 1…クライアント計算機
- 2…キーボード
- 3…情報提供サーバ
- 4…文書
- 5…ネットワーク
- 6…検索サーバ
- 7…文書情報登録部
- 8…文書情報保存部
- 9…差分抽出部
- 10…検索履歴情報データベース
- 11…差分通知部
- 12…検索部
- 13…送信する手段

【図1】



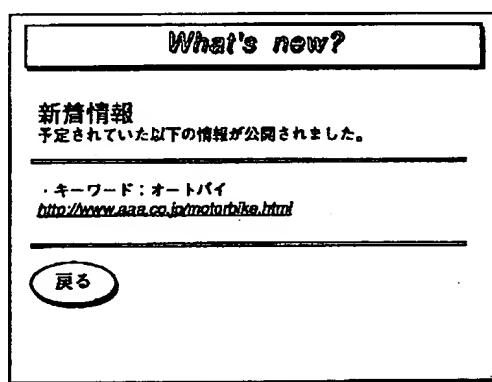
【図2】



【図3】



【図5】



【図4】

